

automatic speech recognition (ASR) and text-independent speaker recognition to provide secure access to services and/or facilities (Kanevsky, col. 5, lines 34-38). Two simultaneous processes occur to verify the identity of a user in order to provide the user access. Both processes utilize user provided answers to randomly system-generated questions. The security system uses two pre-existing databases, a user database 18 of user-specific non-acoustic information and an acoustic user model 20 of user-specific voice samples, to match the user provided answers to the pre-existing user information in the two databases. A successful match indicates verification of the user's identification. The two databases 18 and 20 contain different types of data. Database 18 contains text information that correspond to correct user-specific answers to the random questions. Such answers may include name, address, phone-number, identification number, or the like. Database 20 contains acoustic information in the form of speech samples or voice prints that correspond to correct user-specific verbal answers to the same random questions.

The first process uses the database 18 and involves the user correctly answering randomly generated questions via an iterative process. In the first process, the user calls a central server 22 and identifies himself via his name, for example (Kanevsky, col. 6, lines 4-5). The server 22 submits the user utterance of the user's name to ASR 28 (Kanevsky, col. 6, lines 7-9). The ASR 28 decodes the utterance and submits the decoded name back to the server 22 (Kanevsky, col. 6, lines 9-10). The ASR 28 is a conventional speech recognition device (Kanevsky, col. 13, lines 48-61). The server 22 then accesses the database 18 to find a user-specific database corresponding to the user identifying himself by name above (Kanevsky, col. 6, lines 16-18). It should be clear that the name identified by the ASR 28 is not intended as any form of identity verification. Instead, the name identified by the ASR 28 is used as an identity verification starting point. As a starting point, the identified name is used to access a user-specific database within the database 18. The user-specific database contains the pre-existing user-specific non-acoustic information. This user-specific database is then used as a basis to perform the identity verification.

Next, utilizing the user-specific information from the identified user-specific database, the server 22 generates a random question (or multiple random questions) for the user to answer (Kanevsky, col. 6, lines 25-27). The user answers the question(s) which is sent back to the server 22. The server 22 then sends the user answers to the ASR 28 to be decoded (Kanevsky, col. 6, lines 27-35). After decoding the answer, the ASR 28 passes the decoded answer to the semantic

analyzer 40 to determine if the answer is correct, or not, in accordance with the user-specific information in the user-specific database (Kanevsky, col. 6, lines 35-39). The result of the semantic analyzer 40 is sent to a score estimator 44 where a partial score associated with the answer received from the user is generated (Kanevsky, col. 6, lines 39-42). The partial score is sent to the server 22. The question and answer process between the server 22 and the user may continue for as many iterations as are desired to substantially ensure that the user answering the questions is the user associated with the identified user-specific database within the database 18 (Kanevsky, col. 6, lines 61-64).

The second process uses the acoustic user model 20. Simultaneous with the question and answer session associated with the first process, the server 22 sends a user voice sample from the answers uttered by the user to a text-independent speaker recognition module 52 (Kanevsky, col. 6, line 66 to col. 7, line 4). The text-independent speaker recognition module 52 compares the user voice sample obtained from the user's answer to user-specific acoustic samples found in a user-specific acoustic model within the acoustic user model 20 (Kanevsky, col. 7, lines 5-12). The user-specific acoustic model corresponds to the user identified in the beginning of the first process. *The user voice sample is processed against the user-specific acoustic model by the speaker recognition module 52*, and the results are sent to the score estimator 44 to generate another partial score (Kanevsky, col. 7, lines 12-15). Based on a comparison of a combination of the partial score from the first process (measures correct responses to questions) and the partial score from the second process (speaker verification based on similar voice samples) versus a predetermined threshold, the server 22 decides whether or not to permit the user access to the service/facility (Kanevsky, col. 7, lines 15-20). Clearly, from the description above, the database 18 and the acoustic user module 20 are not speech recognition or speaker recognition/identification applications, they are merely databases. **Further, the acoustic user model 20 is utilized by the text-independent speaker recognition module 52 as a template of known acoustic answers by which the second process of speaker/identification is performed. The acoustic user model 20 is not utilized by the ASR 28.** The ASR 28 utilizes the database 18 as a known template of text answers by which the first process of speaker identification/verification is performed.

Within the Office Action, it is stated that Kanevsky teaches a system that collects voice samples from a user and uses the collected voice samples to modify the ASR 28. To support this

assertion, column 8, lines 48-50 of Kanevsky is cited. The Applicant respectfully disagrees with this interpretation of this passage. In lines 48-50 of column 8, Kanevsky teaches "...the system collects voice samples from the caller's answers to the plurality of questions and builds a user voice model (e.g. user model 20) therefrom." **As discussed above in relation to the second process, the acoustic user model 20 is used by the text-independent speaker recognition module 52, not by the ASR 28.** The ASR 28 is used to decode the verbal answers provided by the user to questions generated by the server 22 in the first process. It is the semantic analyzer 40 that determines if the decoded answer from the ASR 28 matches pre-existing text information stored within the database 18. By adding voice samples to the user acoustic model 20, as described in the cited passage, the number of pre-existing voice samples within the acoustic user model 20 is increased, but there is no impact on the database 18 and therefore no impact on the ASR 28. By increasing the usable number of voice samples in the acoustic user model 20, there is an improved likelihood that a user utterance in conjunction with the second process is matched by the text-independent speaker recognition module 52. Therefore, by adding voice samples to the user acoustic model 20, the efficiency of the text-independent speaker recognition module 52 is improved. This improves the opportunity to increase the partial score in the second process. However, this has no bearing on the first process other than possibly reducing the number of questions necessary to reach a partial score sufficient to gain access to the system, since the other partial score of the second process is possibly increased (Kanevsky, col. 8, lines 51-55).

In contrast to the teachings of Kanevsky, the speech recognition system of the present invention provides for speaker-specific acoustic models to be used in the **speech recognition process**. Multiple users can access the same application, each user having an individualized speaker-specific acoustic model which is stored, retrieved from storage, and modified according to samples of the specific speaker's speech. By utilizing speaker-specific models which are uniquely tailored to the individual user, the speech recognition system of the present invention greatly improves the accuracy of speech recognition over that of a generalized speech recognition system, such as the ASR 28 taught by Kanevsky. Kanevsky teaches a system to modify a **speaker identification process**. Kanevsky does not teach a system to modify a **speech recognition system**.

Within the Office Action, it is acknowledged that Kanevsky teaches a method of modifying a speaker identification system. The Applicant agrees. However, it is further stated

within the Office Action that the present application also teaches a method to modify a speaker identification system. The Applicant respectfully disagrees. To support this assertion, the Examiner states that the independent claims of the present application claim obtaining an identification of a speaker by the speaker's name, and obtaining a speech sample of the speaker during remote session to modify speaker specific model. The Applicant respectfully points out that the independent claims, and specifically independent Claim 1, teaches "obtaining an identification of a speaker." There is no limitation as to how the identification is obtained. In fact, the specification states that any number of conventional identification techniques can be used including prompting the speaker to speak his name, entering a personal identification number, entering an account number, automatically receiving the speaker's caller ID for a telephone call, and utilizing voice identification techniques (Specification, page 5, lines 4-9). Independent claim 1 further teaches "obtaining a sample of a speaker's speech during a first remote session", "recognizing the speaker's speech utilizing the speech recognition system during the first remote session", "modifying the speech recognition system according to the sample thereby forming a speaker-specific modified speech recognition system", and "storing a representation of the speaker-specific modified speech recognition system in association with the identification of the speaker." Clearly, the independent claims of the present application use the identification of the speaker as a tag, the identification is not used to permit or deny access to the system. In contrast, and as acknowledged in the Office Action, Kanevsky teaches a system that determines whether or not to permit a user access to a system. Using identification as a tag and testing the credibility of the identification are not the same, and do not teach the same limitation as stated in the Office Action. Therefore, the assertion that both Kanevsky and the present application teach the same limitation is not valid.

The independent Claim 1 is directed to a method of adapting a speech recognition system. The method of Claim 1 includes the steps of obtaining an identification of a speaker, obtaining a sample of a speaker's speech during a first remote session, recognizing the speaker's speech utilizing the speech recognition system during the first remote session, **modifying the speech recognition system** according to the sample thereby forming a speaker-specific modified speech recognition system, storing a representation of the speaker-specific modified speech recognition system in association with the identification of the speaker, and using the representation of the speaker-specific modified speech recognition system to recognize speech during a subsequent

remote session with the speaker. As discussed above, Kanevsky teaches a system to modify a speaker identification process. Kanevsky does not teach a system to modify a speech recognition system. For at least these reasons, Claim 1 is allowable over the teachings of Kanevsky.

Claims 2-10 and 12-16 are each dependent upon the independent Claim 1. As discussed above, the independent Claim 1 is allowable over the teachings of Kanevsky. Accordingly, Claims 2-10 and 12-16 are each also allowable as being dependent upon an allowable base claim.

Within the Office Action, Claims 17-26, 28-40, and 42-58 have been rejected as having similar limitations as Claims 1-10 and 12-16. The Applicant respectfully traverses this rejection for at least the same reasons as discussed above pertaining to Claims 1-10 and 12-16.

For the reasons given above, Applicant respectfully submits that all of the remaining claims are in a condition for allowance, and allowance at an early date would be appreciated. Should the Examiner have any questions or comments, he is encouraged to call the undersigned at (408) 530-9700 to discuss the same so that any outstanding issues can be expeditiously resolved.

Respectfully submitted,
HAVERSTOCK & OWENS LLP

Dated: 5-21-02

By: Thomas B. Haverstock
Thomas B. Haverstock
Reg. No. 32,571
Attorneys for Applicants

CERTIFICATE OF MAILING (37 CFR § 1.8(a));

I hereby certify that this paper (along with any referred to as being attached or enclosed) is being deposited with the U.S. Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to the: Assistant Commissioner for Patents, Washington D.C. 20231

HAVERSTOCK & OWENS LLP

Date: 5-21-02 By: Thomas B. Haverstock